

# Multidimensional item response theory to assess psychological properties of GHQ-12 in school children parents

Elham Haem<sup>1</sup>, Marziyeh Doostfateme<sup>1</sup>

<sup>1</sup> Department of Biostatistics, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran

**Corresponding author:** Marziyeh Doostfateme<sup>1</sup>

Mailing address: Department of Biostatistics, Medical school, Emam Hossein square, Shiraz, Iran

Office telephone number: +987130283015

Fax number: +987132349930

Email address: m.fatemi82@gmail.com

## Abstract

**Background:** Multidimensional item response theory (MIRT) model provides an ideal foundation to assess psychological properties of a questionnaire designed with multidimensional structure. This study aims to present the first use of MIRT models to investigate psychological properties of GHQ-12.

**Methods:** A total of 1104 school children parents completed the Persian version of GHQ-12 questionnaire. A MIRT model was applied to model observed scores for each GHQ-12 item as a function of the subject's latent traits while taking into account the correlation between dimensions of the questionnaire. The goodness of fit indices were reported and items fit were assessed. Individual items were described in detail through item characteristic curves and the amount of information carried by different items were presented using information curves.

**Results:** The MIRT analysis of the 12 item GHQ successfully demonstrated two factor structure corresponding to psychological distress and social dysfunction. The model showed all items were fitted adequately. Items varied in their discrimination ranged from 0.86 to 2.35 and 1.18 to 2.41 for psychological distress and social dysfunction, respectively. Moreover, item 8 and item 2 provided the least information in psychological distress and social dysfunction dimensions, respectively.

**Conclusions:** The developed framework to evaluate psychological properties of GHQ-12 can be a suitable alternative to traditional approaches and also unidimensional IRT models which its use have been restricted due to multidimensional structure of the questionnaire.

**Keywords:** Multidimensional item response theory, GHQ-12, psychological properties, school children parents

## Background

The general health questionnaire (GHQ) is a self-report measure of minor psychiatric morbidity that has been widely used, since its development by Goldberg in 1972 [1]. The original instrument consists of 60 items, but different shorter versions including GHQ-30, GHQ-28 and GHQ-12, have also been adapted and validated in different studies [2]. The 12-items version of the questionnaire, GHQ-12,6 was used broadly due to its relatively good psychometric properties and its brevity [3,4]. Further, the GHQ-12 is recommended by world health organization (WHO) as a well-validated and standard psychiatric screening instrument [5].

The GHQ-12 consists of 12 items, each of which is rated on a four-point scale, typically worded: less than usual, no more than usual, rather more than usual, or much more than usual. The two most commonly used scoring methods are bi-modal (0-0-1-1) and Likert scoring styles (0-1-2-3) [6].

Since the GHQ-12 exhibits considerable appeal as a quick and well-documented screening tool, it was translated into different languages to study its reliability and validity and explore its psychometric properties in various population and countries [6-12]. For the first time, the Persian version of the questionnaire was prepared and its psychometric properties were assessed by Montazeri et al [13]. Since then, several studies to assess its applicability among university students and Iranian elder population were conducted [14,5].

The questionnaire was designed as a unidimensional scale to capture a single trait and some empirical studies supported this assumption [15,16]. However, more frequently studies have clearly revealed the existence of two or three factor solutions [12]. Most of the studies yielded a two factor solution named “anxiety/depression” and “social dysfunction” [17-19,7,20-22]. Some studies, however, revealed a third factor expressing “loss of confidence” [23-25]. For Persian version of the questionnaire, a two factor model was the best explanation of the Iranian sample[13].

Traditionally, classical test theory (CTT) including construct validity, reproducibility and sensitivity to change were used to assess psychometric properties of questionnaires [26]. Furthermore, confirmatory factor analysis (CFA) as a common method can be used to evaluate hypothesis about the dimensionality of questionnaires [27]. Although, CTT and CFA are popular methods, they do not consider important aspects of a questionnaire such as item difficulty, item discrimination and ordering of responses categories [26].

The item response theory (IRT) provides a more detailed assessment of a questionnaire’s items. This theory, also known as the latent response theory attempts to explain the relationship between an individual responses to items on the questionnaire and the latent trait [28,29]. It establishes a link between the properties of items on a questionnaire, individuals responding to these items and the underlying trait being measured.

Despite IRT benefits, this approach seldom is used to investigate properties of a questionnaire. Depaoli et al. [27] reviewed articles published since 2005 across the journals that incorporated scale development or assessment in the field of Health Psychology. Of 126 articles only five used IRT-based models and the remaining used CTT methods [27].

Most conducted studies to investigate psychometric properties Of GHQ-12 used CTT methods, exploratory factor analysis and confirmatory factor analysis. The use of unidimensional IRT models to assess factorial structure of GHQ-12 is really sparse. When questionnaires comprise

multiple dimensions, the utility of unidimensional IRT is largely restricted. An improved version of IRT models named multidimensional IRT (MIRT) models take into account multiple latent traits simultaneously and correlation amongst latent traits were considered. To the best of our knowledge, no study applied MIRT models for assessing psychometric properties and dimensionality assessment of GHQ-12.

Further, it appears that there is no reported study on the psychiatric morbidity of parents with schoolchildren measured by GHQ-12 in Iranian population. Whereas children's quality of life is one of the important and complementary outcomes in clinical studies, several studies have focused on this subject [30-32]. On the other hand, health-related quality of life in children is strongly influenced by the mental health of their parents. Therefore, it is crucial to evaluate the parents' psychiatric morbidity in a population.

The present study aimed to use a MIRT model to investigate properties of the questionnaire with more detail. Since the Persian version of GHQ-12 [13] was described well with two factors, a MIRT model including two factors namely psychological distress and social dysfunction was applied. Individual items were described in detail through item characteristic curves and item information curves.

## **Methods**

### **Participants and instrument**

The Persian version of the GHQ-12 translated and validated previously in Iran [13], was filled in by 1104 parents of Iranian secondary school adolescents aged 13-18. A two-stage cluster random sampling technique was used to select participants randomly. At the first stage, four schools were selected at random from 60 secondary schools in each of four educational districts in Shiraz, southern Iran. Afterwards, two classes from each school were chosen through a simple random sampling and all parents of the students in the chosen classes were considered as the study population in the second stage. The students took home the informed consent forms and the questionnaires for their parents and then the filled questionnaires were returned to the schools. The ethics committee of Shiraz University of Medical Sciences approved the study.

The GHQ-12 includes of 12 ordered categorical questions or items are rated between four categories 0, 1, 2 and 3 indicating less than usual, no more than usual, rather more than usual, or much more than usual, respectively. The GHQ-12 scoring protocol has reversed-scored items such that the higher scores show better psychological health state which model was fitted accordingly.

### **Multidimensional item response theory**

IRT models assumes that there is only one latent variable  $\theta$  to explain the relationship between latent traits and observed responses. However, MIRT as an extension of IRT models attempts to explain an item response according to an individual's standing multiple latent dimensions[33]. There are several forms of IRT models that have been used for ordered categorical data including rating scale model, partial credit model, generalized partial credit model (GPCM), and graded response model (GRM) [29]. The most common IRT-based approach for multiple-response questionnaires in patient-reported outcome studies has been GRM [28]. In this study, a

multidimensional extension of GRM was used to describe the probability of a given score as a function of two latent variables. Since a two factor structure was explored for Persian version of GHQ-12 [13], a multidimensional GRM with two factor named “psychological distress” and “social dysfunction” considered in this study.

The functional form of the multidimensional GRM is given by:

$$P(Y_{ij} \geq K | \theta_i = \theta) = \frac{1}{1 + \exp[-(a_j^T \theta + c_{jk})]} \quad (1)$$

$$P(Y_{ij} = K | \theta_i = \theta) = P(Y_{ij} \geq K | \theta_i = \theta) - P(Y_{ij} \geq K + 1 | \theta_i = \theta) \quad (2)$$

Where  $P(Y_{ij} \geq K | \theta_i = \theta)$  is the probability that observed scores for item j and subject i given the ability on latent trait  $\theta_i$  obtain a score greater or equal to k, with k=0 to 3. In this equation,  $a_j$  and  $c_{jk}$  denote, respectively, the item discrimination and intercept, where intercepts are ordered and one less than the number of response categories for each item. A high discrimination value shows that an item is able to differentiate between subjects at different latent trait levels. The intercept,  $c_{jk}$  can be transformed into a difficulty parameter,  $b_{jk}$  through the following formula,

$$b_{jk} = \frac{-c_{jk}}{a_j} \quad (3)$$

Where a low value for difficulty parameter indicates an easy item and a high difficulty indicates a difficult item. Further, in Eq (1), latent traits are distributed normally,  $\theta_i \sim N(0, \Omega)$ , where  $\Omega$  is the covariance matrix for individual i's latent traits. The correlation between dimensions takes into account in the multidimensional GRM model through  $\Omega$  [27].

## Statistical analysis

All analysis were performed in the R programming environment with the multidimensional item response theory (mirt) package [34]. Item characteristic curves (ICC) were provided to describe visually the probability of each score in each item. Furthermore, item information curves were included to investigate which items of GHQ-12 carry the most information to detect psychiatric morbidity of parents. Information content of items were calculated using Fisher information which is formulated as minus the expectation of the second derivative of the log-likelihood of the model [28]. To evaluate the item fit, the generalized Orlando and Thissen's S-X<sup>2</sup> index for polytomous data was used [35], comparing the observed and expected response frequencies under the estimated MIRT model. Eventually items with S-X<sup>2</sup> p-value<0.001 were considered poorly fitted [36].

## Result

In this study, there were 13248 observations from 1104 parents of school children. The distributions of observed responses of items for psychological distress and social dysfunction dimensions are shown in Fig. 1. The frequency of ordinal items showed a diverse pattern in two dimensions. In psychological distress dimension, most items being skewed toward high scores (2 and more) indicating the better psychological health state while items of social dysfunction were more symmetrically distributed.

An MIRT model with two factor was fitted on the GHQ-12 data set .Table 1 summarizes the goodness of fit of the model representing all indices satisfies the cut-off values for a good fit. Further, item specific parameters were estimated successfully. Table 2 provides estimation of item discrimination and item difficulty parameters and their standard error for two dimensions. For all items in the two dimension, discrimination estimates ranged from 0.86 to 2.41, indicating that all items discriminated between low and high of levels of GHQ-12 latent traits (or psychological health state) of parents very well.

Fig 2. Shows examples of the obtained ICCs for three items. This figure indicates that a person with better psychological health state (higher latent trait, the latent trait is either psychological distress or social dysfunction) has a higher probability of increased scores for each item. The lowest slope of 0.86 for face up to problems (item 8) indicates a lower discrimination power in psychological distress of parents. In other words, a large increment in health state just yields a small increment in the probability for the score on this item. However, the high slope parameter of 2.41 and 2.35 for feeling unhappy and depressed (item 9) and losing confidence (item 10) indicating a higher discrimination power in social dysfunction and psychological distress latent traits, respectively. For all items, when psychological health state score increases, the probability of a 0 score decreases.

Fig 3. presents item information curves for all items of psychological distress and social dysfunction dimensions, separately. Item information curves as a function of latent traits, indicate which item carry the most information and also where on the latent trait they are most informative. The information content carried by items are different. In social dysfunction, feeling unhappy and depressed (item9) was the most informative over the moderate range of latent trait while lost much sleep (item2) is the least informative over a broad range of the latent trait. Losing confidence (item 10) and Thinking of self as worthless (item 11) carried the most information on the moderate latent trait. However, face up to problems (8) carried little to almost no information in this study.

Table 3 presents full results for item fit statistics. Based on  $S-X^2$  p-value, all the items fit the GHQ-12 questionnaire properly.

Fig1. Distributions of observed item responses (0= much more than usual, 1= rather more than usual, 2= no more than usual, 3= less than usual) for each dimension. The name of items are provided in Table 2.

Fig2. Item characteristic curves showing the probability for each individual score within each category of items for three example of items.

Fig3. Item information curves for items of psychological distress and social dysfunction dimensions.

## Discussion

The present study is the first to apply MIRT model to evaluate psychometric properties of GHQ-12 questionnaire. This study included 1104 parents of school children to measure their minor psychiatric morbidity. Since maternal and paternal psychological health affect children's development and health during school, assessment of their psychiatric morbidity is essential.

The analysis of questionnaires and assessing their psychometric properties through CTT approach focusing on summated scores disregards the underling nature of the data. Traditionally, CFA analysis has been used widely to assess the dimensionality or underlying latent variable structure of a questionnaire.

An IRT model provides some advantages over CFA to assess the dimensionality of a questionnaire. It provides a deeper insight into the measurement properties of a questionnaire and its items. In this approach, ICC curves presents the power of discrimination and difficulty of individual items. Further, item information functions are obtained through IRT models, estimate the precision and reliability of individual items independent of other items on the questionnaire. In addition, item information curves indicate the content of information carried by individual items. As a result, a subset of items can be selected and a reduced questionnaire developed by omitting uninformative items. However, these indices of items and curves are not available in CFA. It only uses the factor loading to show the relationship between items and latent traits and test hypotheses about the dimensionality of the questionnaires [37].

Notwithstanding the advantages of IRT over CFA, it suffers from one limitation which is the need for large samples. A summary of the recommended sample sizes for various IRT models are provided by Yen and Fitzpatrick [38]. MIRT as an extension of IRT approach model multiple dimensions simultaneously to take into account the correlation amongst dimensions. Since these correlation parameters are estimated amongst dimensions, MIRT models need larger sample compared to IRT models. In this study, an enough large sample employed to obtain stable parameter estimates in the MIRT model.

In the present study, the MIRT model with two factor showed an adequate fit for all indices. Our findings corroborate other studies that reported two dimensional structure including psychological distress and social dysfunction although they used CTT and CFA [14,13,5]. IRT- based models are really sparse in order to assess psychometric properties of GHQ-12. Smith et al [39], applied a Rasch model and CFA to the 12-item GHQ and identified 6 misfitting items. In the mentioned study, they focused more on differential item functioning by age, gender and treatment aims. However, the discrimination and difficulty parameters, ICC and information curves were not reported [39]. Our findings highlight no misfitting items which are not in line with the mentioned study. This inconsistency may be explained by the difference between MIRT models, considering correlation amongst dimensions, and unidimensional IRT models. Further, in our study, a graded response model was used through the MIRT model while Smith et al [39] applied a Rasch model in the IRT approach. Since graded response models have fewer assumptions compared to Rasch models, they are more flexible and likely to fit data generated from patient reported outcomes [40].

A benefit of IRT-based models is the amount of item information calculated based on item characteristic curves. They provide the relative contribution of different items to total information across different regions along of the latent trait. Consequently, item information curves play a significant role in describing items, optimal selection of the most informative of subset of items and comparing efficiency between different tests [41,28]. In psychological distress dimension, two items including face up to problems (item 8) and capable of making decision (item 4), were found to have the least information. Furthermore, in social dysfunction dimension, lost much sleep (item 2) included lower information in a broad range of the latent trait compared to other items. Hence, a subset of more informative items can be selected and a shortened version of GHQ-12 developed.

## **Conclusion**

Based on GHQ-12 data from school children parents, a MIRT model with two factors namely psychological distress and social dysfunction was successfully developed to examine the psychometric properties of the questionnaire. Additionally, item fit statistics assessed individual

items. Further, information curves described the amount of information carried by individual items. MIRT models can be adapted as a powerful tool to examine the psychometric properties of questionnaires designed with an intentional multidimensional structure. It is hope that, published articles on MIRT models stimulates its increased use within the health psychology field.

## References

1. Goldberg DP, Blackwell B (1970) Psychiatric illness in general practice: a detailed study using a new method of case identification. *Br med J* 2 (5707):439-443
2. Goldberg DP, Hillier VF (2009) A scaled version of the General Health Questionnaire. *Psychological Medicine* 9 (1):139-145. doi:10.1017/S0033291700021644
3. Romppel M, Braehler E, Roth M, Glaesmer H (2013) What is the General Health Questionnaire-12 assessing?: Dimensionality and psychometric properties of the General Health Questionnaire-12 in a large scale German population sample. *Comprehensive Psychiatry* 54 (4):406-413. doi:https://doi.org/10.1016/j.comppsy.2012.10.010
4. Salama-Younes M, Montazeri A, Ismail A, Roncin C (2009) Factor structure and internal consistency of the 12-item General Health Questionnaire (GHQ-12) and the Subjective Vitality Scale (VS), and the relationship between them: a study from France. *Health Qual Life Outcomes* 7:22. doi:10.1186/1477-7525-7-22
5. Rahmati Najarkolaei F, Raiisi F, Rahnema P, Gholami Fesharaki M, Zamani O, Jafari MR, Montazeri A (2014) Factor structure of the Iranian version of 12-item general health questionnaire. *Iran Red Crescent Med J* 16 (9):e11794-e11794. doi:10.5812/ircmj.11794
6. Liang Y, Wang L, Yin X (2016) The factor structure of the 12-item general health questionnaire (GHQ-12) in young Chinese civil servants. *Health Qual Life Outcomes* 14 (1):136-136. doi:10.1186/s12955-016-0539-y
7. Politi PL, Piccinelli M, Wilkinson G (1994) Reliability, validity and factor structure of the 12-item General Health Questionnaire among young males in Italy. *Acta Psychiatrica Scandinavica* 90 (6):432-437. doi:10.1111/j.1600-0447.1994.tb01620.x
8. Daradkeh TK, Ghubash R, el-Rufaie OE (2001) Reliability, validity, and factor structure of the Arabic version of the 12-item General Health Questionnaire. *Psychol Rep* 89 (1):85-94. doi:10.2466/pr0.2001.89.1.85
9. Quek KF, Low WY, Razack AH, Loh CS (2001) Reliability and validity of the General Health Questionnaire (GHQ-12) among urological patients: a Malaysian study. *Psychiatry Clin Neurosci* 55 (5):509-513. doi:10.1046/j.1440-1819.2001.00897.x
10. Kilic C, Rezaki M, Rezaki B, Kaplan I, Ozgen G, Sağduyu A, Oztürk MO (1997) General Health Questionnaire (GHQ12 & GHQ28): psychometric properties and factor structure of the scales in a Turkish primary care sample. *Soc Psychiatry Psychiatr Epidemiol* 32 (6):327-331. doi:10.1007/bf00805437
11. Chan DW (1993) The Chinese General Health Questionnaire in a psychiatric setting: the development of the Chinese scaled version. *Soc Psychiatry Psychiatr Epidemiol* 28 (3):124-129. doi:10.1007/bf00801742
12. Sánchez-López MdP, Dresch V (2008) The 12-Item General Health Questionnaire (GHQ-12): reliability, external validity and factor structure in the Spanish population. *Psicothema* 20 (4):839-843
13. Montazeri A, Harirchi AM, Shariati M, Garmaroudi G, Ebadi M, Fateh A (2003) The 12-item General Health Questionnaire (GHQ-12): translation and validation study of the Iranian version. *Health Qual Life Outcomes* 1:66-66. doi:10.1186/1477-7525-1-66
14. Namjoo S, Shaghghi A, Sarbaksh P, Allahverdipour H, Pakpour AH (2017) Psychometric properties of the General Health Questionnaire (GHQ-12) to be applied for the Iranian elder population. *Aging & Mental Health* 21 (10):1047-1051. doi:10.1080/13607863.2016.1196337
15. Fernandes HM, Vasconcelos-Raposo J (2013) Factorial validity and invariance of the GHQ-12 among clinical and nonclinical samples. *Assessment* 20 (2):219-229. doi:10.1177/1073191112465768
16. Hahn D, Reuter K, Härter M (2006) Screening for affective and anxiety disorders in medical patients - comparison of HADS, GHQ-12 and Brief-PHQ. *Psychosoc Med* 3:Doc09
17. Vanheule S, Bogaerts S (2005) The factorial structure of the GHQ-12. *Stress and Health: Journal of the International Society for the Investigation of Stress* 21 (4):217-222. doi:doi.org/10.1002/smi.1058

18. Toyabe S-i, Shioiri T, Kobayashi K, Kuwabara H, Koizumi M, Endo T, Ito M, Honma H, Fukushima N, Someya T (2007) Factor structure of the General Health Questionnaire (GHQ-12) in subjects who had suffered from the 2004 Niigata-Chuetsu Earthquake in Japan: a community-based study. *BMC Public Health* 7 (1):175. doi:10.1186/1471-2458-7-175
19. Schmitz N, Kruse J, Tress W (2001) Improving screening for mental disorders in the primary care setting by combining the GHQ-12 and SCL-90-R subscales. *Comprehensive psychiatry* 42 (2):166-173. doi:10.1053/comp.2001.19751
20. Picardi A, Abeni D, Pasquini P (2001) Assessing psychological distress in patients with skin diseases: reliability, validity and factor structure of the GHQ-12. *J Eur Acad Dermatol Venereol* 15 (5):410-417. doi:10.1046/j.1468-3083.2001.00336.x
21. Kalliath TJ, O'Driscoll MP, Brough P (2004) A confirmatory factor analysis of the General Health Questionnaire-12. *Stress and Health* 20 (1):11-20. doi:10.1002/smi.993
22. Gureje O (1991) Reliability and the factor structure of the Yoruba version of the 12-item General Health Questionnaire. *Acta Psychiatrica Scandinavica* 84 (2):125-129. doi:10.1111/j.1600-0447.1991.tb03115.x
23. Campbell A, Walker J, Farrell G (2003) Confirmatory factor analysis of the GHQ-12: can I see that again? *Aust N Z J Psychiatry* 37 (4):475-483. doi:10.1046/j.1440-1614.2003.01208.x
24. French DJ, Tait RJ (2004) Measurement invariance in the General Health Questionnaire-12 in young Australian adolescents. *Eur Child Adolesc Psychiatry* 13 (1):1-7. doi:10.1007/s00787-004-0345-7
25. Shevlin M, Adamson G (2005) Alternative factor models and factorial invariance of the GHQ-12: a large sample analysis using confirmatory factor analysis. *Psychol Assess* 17 (2):231-236. doi:10.1037/1040-3590.17.2.231
26. Goetz C, Ecosse E, Rat A-C, Pouchot J, Coste J, Guillemin F (2011) Measurement properties of the osteoarthritis of knee and hip quality of life OAKHQOL questionnaire: an item response theory analysis. *Rheumatology (Oxford)* 50 (3):500-505. doi:10.1093/rheumatology/keq357
27. Depaoli S, Tiemensma J, Felt JM (2018) Assessment of health surveys: fitting a multidimensional graded response model. *Psychology, Health & Medicine* 23 (sup1):1299-1317. doi:10.1080/13548506.2018.1447136
28. Haem E, Doostfateme M, Firouzabadi N, Ghazanfari N, Karlsson MO (2020) A longitudinal item response model for Aberrant Behavior Checklist (ABC) data from children with autism. *J Pharmacokinet Pharmacodyn*. doi:10.1007/s10928-020-09686-0
29. Doostfateme M, Taghi Ayatollah SM, Jafari P, Jafari P (2016) Power and Sample Size Calculations in Clinical Trials with Patient-Reported Outcomes under Equal and Unequal Group Sizes Based on Graded Response Model: A Simulation Study. *Value Health* 19 (5):639-647. doi:10.1016/j.jval.2016.03.1857
30. Doostfateme M, Ayatollahi SMT, Jafari P, Jafari P (2015) Testing parent dyad interchangeability in the parent proxy-report of PedsQL™ 4.0: a differential item functioning analysis. *Qual Life Res* 24 (8):1939-1947. doi:10.1007/s11136-015-0931-9
31. Christine E, Morse R (2001) Can Parents Rate Their Child's Health-Related Quality of Life? Results of a Systematic Review. *Quality of Life Research* 10 (4):347-357
32. Varni JW, Limbers CA, Burwinkle TM (2007) Parent proxy-report of their children's health-related quality of life: an analysis of 13,878 parents' reliability and validity across age subgroups using the PedsQL 4.0 Generic Core Scales. *Health Qual Life Outcomes* 5:2-2. doi:10.1186/1477-7525-5-2
33. te Marvelde JM, Glas CA, Van Landeghem G, Van Damme J (2006) Application of multidimensional item response theory models to longitudinal data. *Educational and Psychological Measurement* 66 (1):5-34. doi:10.1177/0013164405282490
34. Chalmers RP (2012) mirt: A Multidimensional Item Response Theory Package for the R Environment. *2012 48 (6):29*. doi:10.18637/jss.v048.i06
35. Kang T, Chen TT (2008) Performance of the Generalized S-X2 Item Fit Index for Polytomous IRT Models. *Journal of Educational Measurement* 45 (4):391-406. doi:10.1111/j.1745-3984.2008.00071.x

36. Lameijer CM, van Bruggen SGJ, Haan EJA, Van Deurzen DFP, Van der Elst K, Stouten V, Kaat AJ, Roorda LD, Terwee CB (2020) Graded response model fit, measurement invariance and (comparative) precision of the Dutch-Flemish PROMIS® Upper Extremity V2.0 item bank in patients with upper extremity disorders. *BMC Musculoskelet Disord* 21 (1):170. doi:10.1186/s12891-020-3178-8
37. Osteen P (2010) An Introduction to Using Multidimensional Item Response Theory to Assess Latent Factor Structures. *Journal of the Society for Social Work and Research* 1 (2):66-82. doi:10.5243/jsswr.2010.6
38. Yen W, Fitzpatrick A (2006) Item Response Theory. In.
39. Smith AB, Fallowfield LJ, Fallowfield LJ, Stark DP, Velikova G, Jenkins V (2010) A Rasch and confirmatory factor analysis of the general health questionnaire (GHQ)--12. *Health Qual Life Outcomes* 8:45. doi:10.1186/1477-7525-8-45
40. Nguyen TH, Han HR, Kim MT, Chan KS (2014) An introduction to item response theory for patient-reported outcome measurement. *Patient* 7 (1):23-35
41. Bortolotti SLV, Tezza R, de Andrade DF, Bornia AC, de Sousa Júnior AF (2013) Relevance and advantages of using the item response theory. *Quality & Quantity* 47 (4):2341-2360. doi:10.1007/s11135-012-9684-5